

CURADORIA DIGITAL DE DADOS DE PESQUISA: O CENÁRIO DAS FUNDAÇÕES DE AMPARO À PESQUISA NO BRASIL

Prof. Dra. Laura Vilela Rodrigues Rezende¹, Prof. Dra. Sonia Cruz Riascos²

¹Universidade Federal de Goiás, lauravil.rr@gmail.com.

²Universidade Federal de Pernambuco, sonia.cruzriascos@gmail.com.

RESUMO

A presente pesquisa trata da curadoria digital de dados de pesquisa no âmbito da comunicação científica brasileira, destacando a importância do uso, da preservação e da custódia em acesso aberto. Apresenta-se breve histórico da comunicação científica destacando a invasão de novos recursos eletrônicos e terminologias, com atenção especial ao acesso aberto (*Open Access*) e seus desdobramentos. Discorre-se sobre o cenário brasileiro de mecanismos de registro, disseminação e disponibilização da produção científica, enfatizando a questão da gestão dos dados digitais de pesquisa, em especial aquelas financiadas com recursos públicos. Sabe-se que, ao longo do tempo, a ciência teve como foco a busca por tecnologias que pudessem decifrar e dimensionar o mundo. Como consequência desta busca incessante, tem-se grande volume de dados gerados em laboratórios e institutos de pesquisa abastecendo planilhas infindáveis. Diante disto, pode-se afirmar que a ciência vive hoje um paradigma que vem alterando os rumos da pesquisa científica, a E-Science. Questiona-se, então: Como aproveitar de forma remota e multidisciplinar a quantidade cada vez maior de dados produzidos durante a realização das investigações? De fato, a consulta aos dados de pesquisa ora concluídas, por outras investigações em elaboração, ou ainda, em vias de serem realizadas depende de uma efetiva gestão desses dados para que possam ser armazenados em bases confiáveis garantindo a recuperação ao longo do tempo. Na tentativa de propor ações para gerir dados de pesquisa, tem-se a curadoria digital que se relaciona com o modo de disponibilizá-los em formato digital para acesso futuro. A ideia é tornar esses dados acessíveis, interpretáveis, reutilizáveis e confiáveis para pesquisas ainda não realizadas. Frente ao contexto exposto, o objetivo desta pesquisa foi o de realizar o mapeamento das informações disponíveis sobre investigações científicas financiadas pelas Fundações de Amparo à Pesquisa no Brasil, para identificar o tipo de informação disponível e, em especial, verificar se os dados que fizeram parte dos corpos das pesquisas se encontram disponibilizados para o público nos sites de tais Fundações. Tratou-se de pesquisa quantitativa, com abordagem descritiva, que utilizou-se da técnica do levantamento de dados, ao nível nacional, nos sites das Fundações Brasileiras de Apoio à Pesquisa. As Fundações de Apoio são instituições criadas com a finalidade de dar suporte aos projetos de pesquisa, ensino, extensão e de desenvolvimento institucional, científico e tecnológico, de interesse das Instituições

Federais de Ensino Superior (IFES), como também às Instituições de Pesquisa. São credenciadas ao Ministério da Educação e ao Ministério da Ciência e Tecnologia do Brasil. Ao final da investigação, os resultados obtidos se encontram apresentados com as considerações necessárias acerca do tema para que se possa aperfeiçoar o tratamento de dados digitais e a disponibilização de informações científicas, a exemplo da implementação de sistemas de gestão desses ativos, como os repositórios digitais, no cenário nacional. Tal medida visa favorecer a oferta de subsídios informacionais para o acesso, o uso e a reutilização para o avanço de pesquisas futuras. Ademais, a ideia de curadoria digital se estende além do controle do repositório que arquiva os recursos e envolve a atenção do criador do conteúdo e dos usuários futuros, indicando-se, portanto, a garantia da preservação e da custódia em ambiente digital, para uso e usufruto contínuo de pesquisadores e interessados.

Palavras-chave: Curadoria digital – Dados de pesquisa - Preservação digital - Acesso aberto

INTRODUÇÃO

A partir do surgimento dos meios de comunicação e da difusão do conhecimento científico, - que inicialmente foram alavancados pela invenção da prensa de caracteres móveis de Gutemberg, causando “grande impacto na divulgação de informações”, devido ao fato de que sua “capacidade de multiplicar os exemplares de um livro, - representou um passo importante rumo a uma disseminação melhor e mais rápida das pesquisas” (MEADOWS, 1999, p.4). Seguindo as palavras de Pinheiro e Kuramoto (2012, p. 307) pode-se afirmar que:

A comunicação foi invadida por novos recursos eletrônicos e terminologia não compreendida e definida. Sítios, portais, twitter, facebook e blogs representam e ampliam as possibilidades não somente de comunicar, mas de disseminar e disponibilizar informações na internet, por sua vez adotando técnicas inovadoras da arquitetura da informação. Novos são os mecanismos de registro, disseminação e disponibilização da produção científica – bibliotecas digitais e virtuais, repositórios institucionais e temáticos, portais de periódicos científicos eletrônicos – nos quais protocolos e compartilhamento de dados buscam facilitar as ações da ciência e tecnologia.

Segundo Mueller (2007, p. 137):

A comunicação informal e formal entre pesquisadores e entre esses e a sociedade como um todo foi muito afetada pela tecnologia da informação e comunicação. No Brasil, essa tecnologia começou a ser utilizada nas décadas de 1960 e 1970, mas só se tornou realmente comum nas atividades científicas das universidades a partir da década de 1990. Periódicos eletrônicos, correio eletrônico, acesso à internet, vídeo conferências e até mesmo *blogs* e *podcasts*, entre outras iniciativas, modificaram profundamente a comunicação científica, não apenas as publicações das revistas, mas também maneiras em que se realizam interações pessoais e o processo de produção das revistas científicas. A facilidade de acesso à internet modificou o papel das bibliotecas, ampliando as fontes disponíveis para consulta.

A crise dos periódicos, movimento caracterizado pelo alto custo dos periódicos científicos e sua inacessibilidade aos menos favorecidos, fez surgir um movimento em favor do acesso irrestrito aos produtos de pesquisas científicas através da internet, segundo afirma Costa (2005, p.175) “na verdade, a questão do preço das assinaturas tem sido uma das principais, se não a principal motivação para os movimentos do acesso aberto e dos arquivos abertos”.

Conforme Varela, Barbosa e Guimarães (2009, p. 129):

[...] a iniciativa dos Arquivos Abertos – *Open Archive Initiative* (OAI) [...], surge como uma tentativa na solução para o alto custo das assinaturas de periódicos científicos, a lentidão na disseminação da produção científica, as barreiras impostas pela legislação de direitos autorais; e a necessidade de reduzir o fosso científico, tecnológico e cultural que se instalou entre países ricos, emergentes e pobres, em função das disparidades educacionais e culturais, para as quais contribuem a centralidade, nas regiões mais prósperas, da produção e distribuição do conhecimento, reduzindo a possibilidade do sujeito das regiões periféricas de ampliar suas competências cognitivas e de aprendizagem, situação que contribui significativamente para acirrar o quadro de exclusão e desigualdade social.

A tendência principal da comunicação eletrônica hoje parece ser em direção a uma filosofia aberta, levantando-se questões sobre software aberto, acesso aberto, arquivos abertos, direito autoral, dentre outras.

Ainda em relação a temática do acesso aberto, recentemente a Organização das Nações Unidas para a Educação, a Ciência e a Cultura (UNESCO), em sua publicação: Informe de la consulta latino-americano y del Caribe sobre Acceso Abierto a información e investigación científica: concepto e políticas, resultado da Congregação realizada em Kingston no ano de 2013, convocou diversas autoridades ao nível mundial para debater o referido tema, em virtude da necessidade de se chegar a consensos que permitam a implementação de ações e políticas de Acesso Aberto, em especial na América Latina e Caribe, localidades representadas nessa ocasião pela Congregação. (ORGANIZAÇÃO, 2013, tradução nossa).

Em relação aos benefícios proporcionados pelo movimento de acesso aberto, destaca-se sua função de dinamização e difusão do conhecimento, que apesar de apresentar barreiras físicas e econômicas, nota-se “a possibilidade de elevação da capacidade de produção de novos conhecimentos, os quais por sua vez, serão convertidos em benefícios para a sociedade como um todo.” (LEITE, 2012, p.414-415).

Sabe-se que no século XX, a maioria dados utilizados para a elaboração de teorias científicas hoje estão inacessíveis. As mudanças ocorridas nos formatos e mídias acabou por tornar ilegível um grande volume de dados criados no passado. Hoje, são muitos os desafios em torno da origem remota dos dados além do acesso comunitário a dados distribuídos. O presente estudo apresenta um mapeamento das informações sobre investigações científicas financiadas pelas Fundações de Amparo à Pesquisa no Brasil disponíveis em seus websites. O conceito que fundamenta esta investigação é o de curadoria digital, o qual se estende para além do controle do repositório que arquiva os dados de pesquisa envolvendo também a atenção do criador do conteúdo e dos usuários futuros, indicando, portanto, a garantia da preservação e da custódia em ambiente digital, para uso e usufruto contínuo de pesquisadores e interessados.

REFERENCIAL TEÓRICO

Dados de pesquisa: conceitos e reflexões atuais

Renato Cruz (2011, p.25) afirma que:

Uma visão antiga de ciência e tecnologia, o chamado modelo linear, dizia que, sem preocupação com aplicações práticas, a pesquisa básica gerava conhecimento e depois era transmitido para a pesquisa aplicada e para o desenvolvimento, que transformariam tal conhecimento em produto.

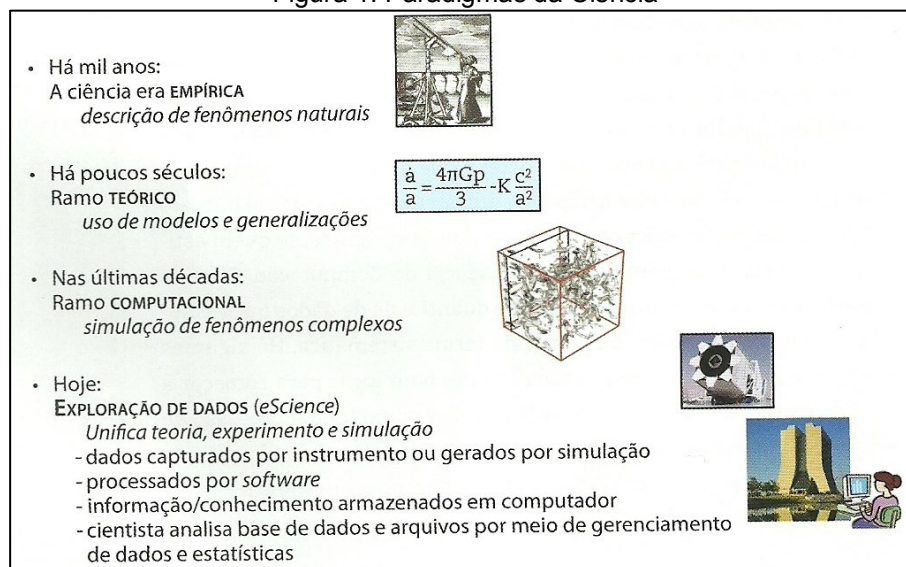
No entanto, ele acrescenta ainda que “há pesquisas que são básicas e, ao mesmo tempo, aplicadas”. Isto quer dizer que geram conhecimento teórico e aplicações práticas. Um aspecto importante no tocante à ciência nos dias atuais é que boa parte dela é alimentada pela tecnologia. Abbot (2011, p.133) afirma que a Tecnologia da Informação (TI) afetou a comunidade científica de duas maneiras. Primeiramente transformando o armazenamento genérico e a computação numa *commodity*. Esta parece ser uma opção razoável para as tarefas científicas que podem ser realizadas por meio de serviços genéricos e centralizados, como a computação em nuvens. A segunda maneira pela qual a TI afetou a comunidade científica, acrescenta o autor, foi por meio da personalização radical. Existe hoje uma rede de serviços de conhecimento distribuídos massivamente cujo desafio é controlá-la. Os cientistas podem criar suas próprias nuvens de computação devido ao acesso pessoal a terabytes de armazenamento e teraflops de computação. É possível visualizar neste cenário uma gama de redes informais de serviços potenciais e uma estreita parceria entre aqueles que criam conhecimento e os que o usam.

Gray (2007 *apud* HEY, TANSLEY, TOLLE, 2011, pág.18, tradução nossa) afirma que:

hoje, o novo modelo de se fazer Ciência prevê a captura de dados por instrumentos ou sua geração através de simulações, seguida por um processamento por software e armazenamento da informação ou conhecimento resultantes em computadores. Os cientistas só veem seus dados numa etapa bastante avançada dessa sequência.

Observando a figura a seguir, ficam evidentes as modificações sofridas pela ciência ao longo do tempo.

Figura 1: Paradigmas da Ciência



Fonte: (GRAY, 2007 *apud* HEY, TANSLEY, TOLLE, 2011, pág.18, tradução nossa)

Cunha (2010) discorre sobre a atual ciência eletrônica afirmando que essa área é, geralmente, composta por bases de dados numéricos e diferentes conjuntos de resultados das pesquisas realizadas nos institutos, faculdades e departamentos.

Junior (*apud* HEY, TANSLEY, TOLLE, 2011, pág.7, tradução nossa) discorre sobre a ciência afirmando que “uma de suas principais características está relacionada com medidas: para estudar um certo fenômeno, medidas objetivas são realizadas sobre amostras ligadas ao fenômeno através de instrumentos (sensores) especializados”. Ele acrescenta ainda que “a evolução da ciência está profundamente ligada à evolução dos instrumentos que permitem a realização de observações”. Ponderando suas considerações ele salienta que “o avanço recente da tecnologia levou a um deslocamento no gargalo para o avanço científico: em vez de a ciência não avançar devido à escassez de dados, hoje em dia ela frequentemente encontra dificuldades em avançar por seu excesso”. Finalizando, “[...] a pesquisa moderna compartilha uma característica comum: a capacidade de gerar e armazenar dados em uma escala sem precedentes e muito além da capacidade humana de análise”.

Ainda nessa publicação, para Junior:

Uma nova área científica nasceu nos últimos dez anos. Como ocorre frequentemente na história da ciência, ela nasceu em diferentes lugares, com diferentes nomes, mas da necessidade comum de se enfrentar o dilúvio de dados: “e-Science”, “ciência orientada por dados” (*data-driven science*), “computação fortemente orientada a dados” (*data-intensive computing*), “ciberinfraestrutura” (*cyberinfrastructure*), “quarto paradigma” (*fourth paradigm*), “dos dados ao conhecimento” (*from data to knowledge*). (HEY, TANSLEY, TOLLE, 2011, tradução nossa).

Curadoria Digital de dados de pesquisa

Na perspectiva de Sayão e Sales (2012), o principal objetivo do conceito de curadoria digital dos dados de pesquisa, configura-se na necessidade de preservar não somente os dados, mas também sua capacidade de transmitir conhecimentos para uso futuro por parte de seus interessados.

Gray (2007 *apud* HEY, TANSLEY, TOLLE, 2011, tradução nossa) afirma que “é preciso se fazer mais em termos de produzir ferramentas para o ciclo completo da pesquisa – captura dos dados e sua curadoria à análise e visualização dos mesmos”. Ele ainda salienta que:

Além da escassez observada em ferramentas para a captura, curadoria e análise de dados, a divulgação dos resultados de pesquisa parte do pressuposto de que todo projeto se encerra em algum momento e não está claro o que ocorre com os dados utilizados. Os dados coletados ao longo de todo o processo da pesquisa geralmente não são curados e nem publicados sistematicamente.

Uma consideração importante e que é consenso quando o assunto se trata de curadoria digital, relaciona-se ao fato de que o conceito é posterior à definição de preservação digital logo, uma vez que esta última se refere às atividades de gerenciamento necessárias para garantir acesso continuado aos materiais digitais pelo tempo que for preciso, pode-se afirmar que não existem ações de curadoria sem preservação digital.

Afirma-se, também, que curadoria digital está relacionada com o modo de tornar dados disponíveis em formato digital para acesso futuro. A ideia é fazer com que esses dados possam ser acessíveis, interpretáveis, reutilizáveis e confiáveis (NEUROTH; STRATHMANN; OBWALD; LUDWIG, 2013).

O *Digital Curation Center* (2014) afirma que a curadoria digital envolve manutenção, preservação e atribuição de valor a dados digitais de pesquisa durante todo o seu ciclo de vida. Ressalta ainda que o gerenciamento efetivo de dados de pesquisa durante todo o seu ciclo diminui os riscos de obsolescência e desvalorização da pesquisa ao longo do tempo.

Lee e Tibbo (2007) afirmam que a curadoria digital é responsável por:

- a) impulsionar a disponibilidade de dados para a comunidade científica;
- b) proporcionar armazenamento redundante;
- c) realizar transformações de dados; e para alguns tipos de materiais, o compromisso de preservação a longo prazo;
- d) disponibilizar dados digitais autênticos para serem reproduzidos e reutilizados;
- e) desenvolvimento de repositórios digitais confiáveis e duráveis;
- f) princípios de criação e captura de metadados de dados multimídia;
- g) uso de padrões abertos para formatos e conversão de arquivos.

Observa-se nas palavras de Corrêa e Bertocchi (2012, p.29) um relato de evolução e contextualização da prática da curadoria digital:

Na medida em que ocorre a expansão da sociedade digitalizada, o termo curadoria passa a ser utilizado para uma diversidade de ações que envolvem organização de dados a partir de critérios ou recortes. Nesse contexto, a “curadoria de informação” assume uma ideia muito mais de organização que de inauguração de uma nova proposta ou visão de mundo.

MATERIAIS E MÉTODOS

As Fundações de Amparo à Pesquisa (FAP) brasileiras são instituições criadas com a finalidade de dar suporte aos projetos de pesquisa, ensino, extensão e de desenvolvimento institucional, científico e tecnológico, de interesse das Instituições Federais de Ensino Superior (IFES), como também às Instituições de Pesquisa. São credenciadas ao Ministério da Educação e ao Ministério da Ciência e Tecnologia do Brasil e compõem o universo deste estudo. O Conselho Nacional das Fundações Estaduais de Amparo à Pesquisa (CONFAP) é uma organização sem fins lucrativos que tem por objetivo melhor articular os interesses das agências estaduais de fomento à pesquisa. Criado oficialmente em 2006, o Conselho agrega FAP de 24 Estados, além do Distrito Federal (BRASÍLIA, 2014).

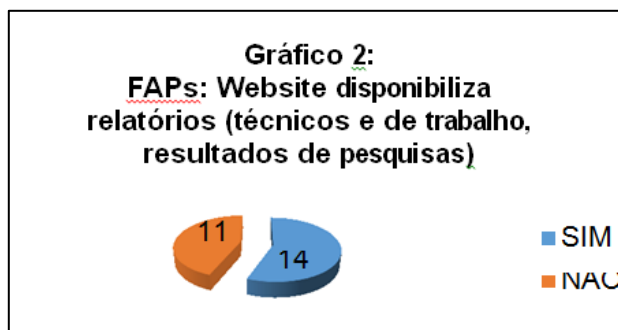
A presente pesquisa se configura com abordagem descritiva, do tipo aplicada e de natureza quantitativa. Utilizou-se da técnica do levantamento de dados, realizado ao nível nacional, nos vinte e cinco (25) websites das Fundações Brasileiras de Apoio à Pesquisa, que configuram a amostra do estudo.

Cenário nacional de acesso a dados de pesquisas financiadas pelas Fundações de Amparo a Pesquisas

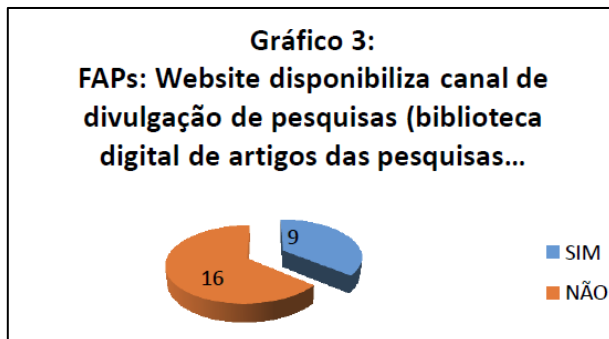
Os resultados de realce estão apresentados a seguir. Inicialmente, verificou-se se os websites das FAP disponibilizam listagens de projetos apoiados. O gráfico 1 ilustra que uma quantidade expressiva de FAP ainda não disponibiliza essas listagens. Tal resultado aponta que 52% das FAP não divulga os projetos realizados, com isso, observa-se a falta de visibilidade dos projetos enquanto estão sendo executados. Alerta-se para a importância de essas informações serem postadas nos websites das Fundações, que todavia não o fazem, para que o acesso aos dados possa mostrar o panorama momentâneo e, com isso, facilitar o desenvolvimento de outros projetos e redirecionar ações similares.



Outro item verificado foi a existência de relatórios técnicos ou de ações, bem como de resultados de pesquisas financiadas. O gráfico 2 ilustra que 56% das FAP disponibiliza este tipo de informação. Todavia, indica-se que as demais FAP (44% da amostra) passem a divulgar os relatórios e resultados, na íntegra com o propósito de apoiar estudos em andamento além de fomentar novas pesquisas.

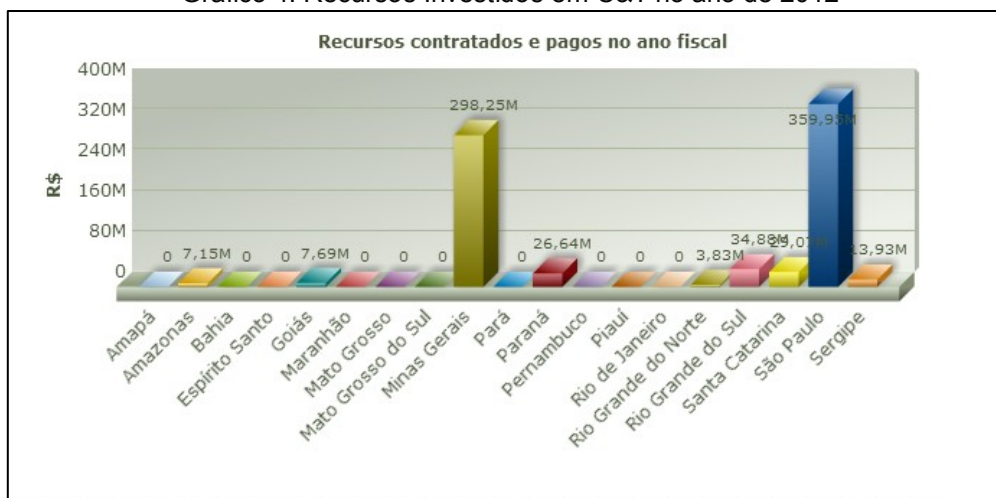


O resultado do levantamento deste item, aponta que 64% das FAP não possuem revistas eletrônicas para a divulgação das pesquisas fomentadas, além de bibliotecas digitais com os artigos que trazem os resultados das investigações financiadas, como pode ser observado no gráfico 3. Trata-se de um quantitativo interessante e pode-se inferir que ocorra incentivo para a visibilidade dos resultados dos estudos empreendidos. Advoga-se, também, pela participação das demais FAP nesse sentido.



A seguir o gráfico 4 aponta os valores de recursos investidos em Ciência e Tecnologia em todas as FAP brasileiras, no ano de 2012. Observa-se maior aporte nos Estados de São Paulo e Minas Gerais; seguidos dos Estados do Rio Grande do Sul e Paraná, isto é, incentivos de relevo nas regiões: sudeste e sul. Sugere-se estudo aprofundado para obtenção de análises detalhadas acerca dos fatores intervenientes em tais resultados.

Gráfico 4: Recursos Investidos em C&T no ano de 2012



O Livro Azul, resultado da 4ª Conferência Nacional de Ciência, Tecnologia e Inovação para o Desenvolvimento Sustentável, sob a coordenação do Ministério da Ciência e Tecnologia, na cidade de Brasília traz como recomendação para o avanço nacional da ciência, no tocante à distribuição de recursos:

É necessário garantir o atendimento às demandas mais sensíveis às assimetrias intra e inter-regionais, concomitantemente ao processo de indução no financiamento da pesquisa científica e tecnológica. Cabe desenvolver mecanismos de descentralização dos recursos para o fortalecimento dos sistemas regionais de C,T&I. Deve ser buscado um melhor entendimento das bases sobre as quais uma agenda de desconcentração possa ser ampliada (2010,p.68).

CONSIDERAÇÕES FINAIS

Sabe-se que, ao longo do tempo, a ciência teve como foco a busca por tecnologias que pudessem decifrar e dimensionar o mundo. Como consequência desta busca incessante, tem-se grande volume de dados gerados em laboratórios e institutos de pesquisa abastecendo planilhas infundáveis. Diante disto, pode-se afirmar que a ciência vive hoje um paradigma que vem alterando os rumos da pesquisa científica, a E-Science. De fato, a consulta aos dados de pesquisa ora concluídas, por outras investigações em elaboração, ou ainda, em vias de serem realizadas depende de uma efetiva gestão desses dados para que possam ser armazenados em bases confiáveis garantindo a recuperação ao longo do tempo. Na tentativa de propor ações para gerir dados de pesquisa, tem-se a curadoria digital que se relaciona com o modo de disponibilizá-los em formato digital para acesso futuro. A ideia é tornar esses dados acessíveis, interpretáveis, reutilizáveis e confiáveis para pesquisas ainda não realizadas.

Sabe-se que ao solicitar fomento para pesquisas científicas, exige-se a apresentação de projetos iniciais e, ao final, na maioria das vezes os resultados obtidos. Uma vez que as agências de fomento já estão de posse destes dados, torna-se mais fácil a implementação de repositórios específicos de dados científicos de pesquisas financiadas.

Ao concluir o presente estudo, observou-se que os resultados obtidos sugerem certa visibilidade dos resultados de pesquisas fomentadas no país com atenção especial voltada para a apresentação de como as agências de fomento têm empregado os recursos disponíveis. Contudo deve-se empreender melhorias quanto ao tratamento de dados digitais, bem como a disponibilização de informações e conteúdos científicos. Sugere-se o planejamento, a implantação e(ou) a implementação de sistemas de gestão desses ativos, por meio de repositórios digitais, institucionais e(ou) temáticos, no âmbito nacional. Essa iniciativa permitiria facilitar o acesso, o uso e a reutilização de informações e de conteúdos de realce para o avanço de pesquisas futuras. Ademais, a ideia de curadoria digital se amplia além do controle do repositório que arquiva os recursos e envolve a atenção do criador do conteúdo e dos usuários futuros, indicando-se, portanto, a garantia da preservação e da custódia em ambiente digital, para uso e usufruto contínuo de pesquisadores e interessados.



REFERÊNCIAS

ABOTT, Mark R. Um novo camino para a ciencia? In: HEY, Tony; TANSLEY, Stewart; TOLLE, Kristin. The Fourth Paradigm: Data-Intensive Scientific Discovery, 2009. Disponível em: <http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_jim_gray_transcript.pdf>. Acesso em: 20 jul. 2014.

CONSELHO NACIONAL DAS FUNDAÇÕES ESTADUAIS DE AMPARO À PESQUISA (CONFAP). Disponível em: < <http://confap.org.br/news/>>. Acesso em: 08 jul. 2014.

CORRÊA, Elisabeth Saad; BERTOCCHI, Daniela. O papel do comunicador num cenário de curadoria algorítmica de informação. In: CORRÊA, Elisabeth Nicolau Saad (Org.). Curadoria digital e o campo da comunicação. São Paulo: ECA/USP, 2012. p.22-29. Disponível em: <http://issuu.com/grupo-ecausp.com/docs/ebook_curadoria_digital_usp/57#download>. Acesso em: 20 mai. 2014.

COSTA, Sely Maria de Souza. O novo papel das tecnologias digitais na comunicação científica. In: MARCONDES, Carlos H. et al. (Org.). Bibliotecas digitais: saberes e práticas. Brasília: IBICT, 2005. p. 167-185.